A Novel Family in *Medicago truncatula* Consisting of More Than 300 Nodule-Specific Genes Coding for Small, Secreted Polypeptides with Conserved Cysteine Motifs[1[w]]

Peter Mergaert[2], Krisztina Nikovics[2], Zsolt Kelemen, Nicolas Maunoury, Danièle Vaubert, Adam Kondorosi, and Eva Kondorosi*

Institut des Sciences du Végétal, Centre National de la Recherche Scientifique, Avenue de la Terrasse, 91198 Gif-sur-Yvette, France

Transcriptome analysis of *Medicago truncatula* nodules has led to the discovery of a gene family named *NCR* (nodule-specific cysteine rich) with more than 300 members. The encoded polypeptides were short (60–90 amino acids), carried a conserved signal peptide, and, except for a conserved cysteine motif, displayed otherwise extensive sequence divergence. Family members were found in pea (*Pisum sativum*), broad bean (*Vicia faba*), white clover (*Trifolium repens*), and *Galega orientalis* but not in other plants, including other legumes, suggesting that the family might be specific for galegoid legumes forming indeterminate nodules. Gene expression of all family members was restricted to nodules except for two, also expressed in mycorrhizal roots. *NCR* genes exhibited distinct temporal and spatial expression patterns in nodules and, thus, were coupled to different stages of development. The signal peptide targeted the polypeptides in the secretory pathway, as shown by green fluorescent protein fusions expressed in onion (*Allium cepa*) epidermal cells. Coregulation of certain *NCR* genes with genes coding for a potentially secreted calmodulin-like protein and for a signal peptide peptidase suggests a concerted action in nodule development. Potential functions of the NCR polypeptides in cell-to-cell signaling and creation of a defense system are discussed.

Plants have evolved symbiotic associations with soil microorganisms to facilitate their mineral nutrition. An example is the specific interaction of different species of the Leguminosae (legumes) with the nitrogen-fixing soil bacteria from the Rhizobiaceae family (rhizobia). This symbiosis leads to the de novo formation of a root organ, the nodule, hosting nitrogen-fixing rhizobia that feed the host plant with ammonium. Another example is the widespread association of plants with fungi from the order of Glomales leading to the formation of arbuscular endomycorrhiza that extends the plant root system and facilitates nutrient uptake. The initial stages of rhizobial and mycorrhizal interactions share certain common molecular mechanisms (Albrecht et al., 1999; Kistner and Parniske, 2002). Because mycorrhizas are more common and ancient, the rhizobial symbiosis might have acquired existing mechanisms from them.

[1] This work was supported in part by "Action Puces à ADN-Centre National de la Recherche Scientifique" (grant), by the Centre National de la Recherche Scientifique-Hungarian Academy of Sciences "Jumelage" program (fellowships to K.N. and Z.K.), and by the Ministère de la Recherche et de la Technologie (fellowship to N.M.).

[2] These authors contributed equally to the paper.

[w] The online version of this article contains Web-only data. The supplemental material is available at www.plantphysiol.org.

* Corresponding author; e-mail eva.kondorosi@isv.cnrs-gif.fr; fax 33–1–69–82–36–95.

Article, publication date, and citation information can be found at www.plantphysiol.org/cgi/doi/10.1104/pp.102.018192.

Two major types of legume nodules are distinguished (Crespi and Gálvez, 2000): the indeterminate type, formed by e.g. *Medicago truncatula*, pea (*Pisum sativum*), broad bean (*Vicia faba*), white clover (*Trifolium repens*), or *Galega orientalis*, and the determinate type, formed by e.g. *Lotus japonicus* or soybean (*Glycine max*). Indeterminate nodules have a complex structure composed of different central tissues surrounded by a cortex (Vasse et al., 1990). The persistent apical meristem is zone I. In zone II, postmeristematic cells gradually differentiate and become infected with rhizobia, encapsulated in a membrane envelope. Interzone II–III is characterized by amyloplast accumulation and major transcriptional changes in both plant and bacterial cells. The proximal zone III is composed of plant cells filled with thousands of nitrogen-fixing rhizobia (bacteroids). In the determinate nodules, cell division occurs only at the early stage of development, and nodules reach their final size by cell elongation. The central tissue of determinate nodules is uniform and contains nitrogen-fixing cells.

To identify genes involved in nodule formation (nodulin genes), various methods have been used such as differential screening, substractive hybridization, differential display, etc. (Crespi and Gálvez, 2000). Using reverse genetics, the function of few genes and their products were studied (Crespi and Gálvez, 2000). Genetic approaches have led to the identification of numerous mutants affected in nodule development, and recently, certain of the corresponding genes were cloned (Schauser et al., 1999; Endre et al., 2002; Stracke et al., 2002).

Technological innovations, such as high-throughput sequencing of expressed sequence tags (ESTs) and DNA arrays, provide new tools for understanding biological processes from a more global viewpoint. Previously, we described the first collection of *M. truncatula* nodule ESTs (Györgyey et al., 2000). At present, 164,441 EST entries originating from 31 cDNA libraries are publicly available in The Institute for Genomic Research (TIGR) *M. truncatula* gene index (MtGI Release 5.0). The ESTs corresponding to transcripts of the same gene are clustered in tentative consensus (TC) sequences producing a set of unique virtual transcripts made of TCs and singletons (only one available EST; Quackenbush et al., 2001). Moreover, the relative abundance of ESTs composing a TC (or singleton) in the different libraries serves as an "electronic northern" for the expression pattern of the genes. Thus, the MtGI can be used to identify nodule-specific genes (Quackenbush et al., 2001).

Here, we describe the discovery of an extremely large gene family from *M. truncatula* that, with exception of the galegoid group of legumes, is absent in other organisms. The encoded polypeptides are characterized by their small size, the conserved Cys motifs, and are probably secreted. All genes of this family exhibited nodule-specific expression, however, with differences in their spatial and temporal expression profiles. Moreover, they were coregulated with calmodulin (CaM)-like and signal peptide peptidase (SPP) genes. Possible functions of the encoded polypeptides in the nodule formation and functioning are discussed.

RESULTS

A Large Gene Family in *M. truncatula*

Previously, we identified 42 *M. truncatula* cDNAs that were induced during nodule development and had no homology to known sequences (Györgyey et al., 2000). Analysis of the putative encoded proteins revealed that 19 of the 42 had similar features and could be classified in the same family based on their small size (about 70 amino acids), the presence of a conserved signal peptide, and conservation of Cys residues at the C-terminal domain (see also below). Due to their expression in the nodule and their Cys content, we named the family NCR (nodule-specific Cys rich). Because the 19 genes derived from a small set of ESTs, it was possible that larger collections may contain additional members of this family. A screen of the TIGR MtGI with successive rounds of BLASTn and TBLASTn searches revealed 311 distinct TCs or singletons belonging to the *NCR* family. Using a similar approach, part of these TCs was found also by Fedorova et al. (2002). A complete list of the *M. truncatula* NCR family members with their accession number, nucleotide, and predicted polypeptide sequences is provided in the supplemental data set (see www.plantphysiol.org).

Genomic Southern blot, using the *NCR001* cDNA as a hybridization probe at low stringency, displayed multiple bands as expected for a multigene family (Fig. 1A). A similarly complex hybridization pattern was obtained with genomic DNA of the tetraploid, cultivated alfalfa, indicating that a comparable large *NCR* family exists also in this species (Fig. 1A).

The alignment of the predicted polypeptide sequences (see supplemental data at www.plantphysiol.org; Fig. 1B) revealed relatively low homologies ranging from 70% to 20% identity at the amino acid level and from 90% to no significant homology at the nucleotide level. Despite the low overall homologies, all members displayed the conserved features of the *NCR* family. All genes coded for small mRNAs of about 400 to 700 nucleotides and polypeptides of 60 to 70 amino acids, except for a few that were somewhat longer (up to 141 amino acids). All carried a hydrophobic amino-terminal domain of 20 to 29 amino acids, predicted with high probabilities by the SignalP program to behave as a signal peptide (Fig. 1B). For most of them, a cleavage site for removal of the signal peptide was predicted. The sequence of the signal peptide domain was relatively well conserved, even in the most distantly related members of the family (Fig. 1B). In contrast, the remaining part of the polypeptides was highly divergent with the exception of conserved Cys with constant spacing between them (Figs. 1B and 2C). The NCR family could be divided in two major groups (groups A and B in Fig. 1B): The first one contained four Cys (C1 and C2, spaced by five amino acids, and C4 and C5, spaced by four amino acids), whereas the second one had two additional conserved Cys (C3 between C2 and C4 with variable spacing and C6, spaced with one amino acid to C5). In addition, hydrophobic residues, spaced one amino acid N terminal to C1, an Asp and a Pro surrounding C2, a basic amino acid (Arg or Lys) preceding and a hydrophobic amino acid after C4, and one or several Pro between C2 (or C3) and C4 were relatively well conserved (see Fig. 1B). In 17 of the 311 polypeptides, the common four-amino acid spacing between C4 and C5 varied. In 13 of the 311 polypeptides, not all of the four Cys residues were found (see supplemental data set at www.plantphysiol.org); however, it should be noted that some of these variations might be due to errors in single-pass sequences of ESTs.

The NCR Family Is Conserved in the Galegoid Group of Legumes and Absent in Other Plants

BLAST searches in the National Center for Biotechnology Information database with several NCR polypeptides led to the identification of sequences that could be grouped in the NCR family. Eight polypeptides derived from pea (Scheres et al., 1990;

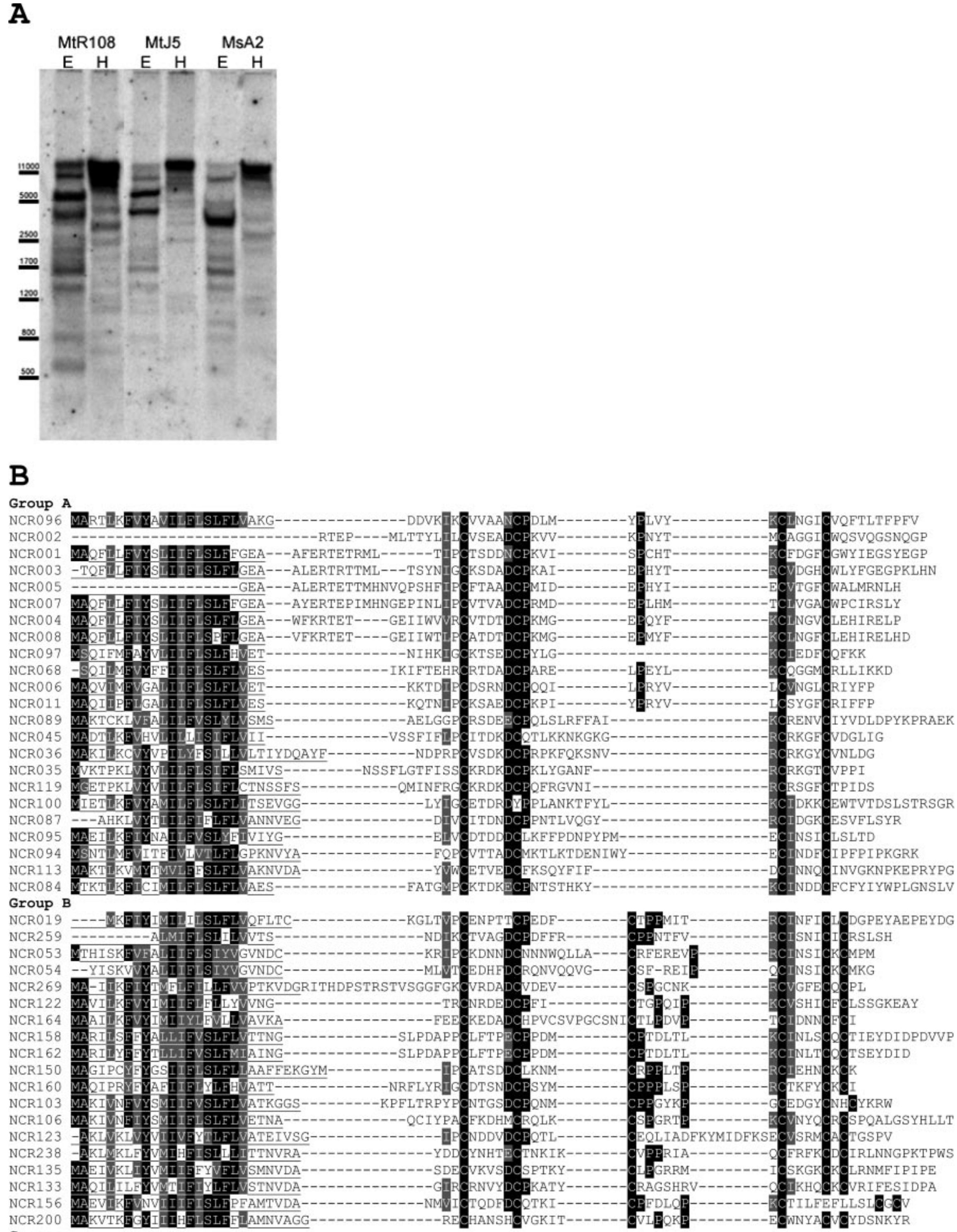

Figure 1. The *NCR* multigene family of *Medicago*. A, Southern hybridization of genomic DNA digested with *EcoRI* (E) or Hind*III* (H) from *M. truncatula* and alfalfa (*Medicago sativa*) with the *NCR001* probe revealing multiple hybridizing bands. MtR108, *M. truncatula* line R108; MtJ5, *M. truncatula* line Jemalong J5; MsA2, alfalfa subsp. *varia* line A2. The Hind*III* digestion was partial. B, Alignment of the 42 most abundantly expressed NCRs (minimum 10 EST hits) forming group A with four and group B with six Cys residues. Identical residues conserved in at least 60% of the sequences are in black, whereas the similar amino acids are in gray background. The signal peptides are underlined. In the consensus sequence, "h" stands for a hydrophobic residue (I, L, V, or M), and "b" stands for a basic residue (R or K). The conserved Cys are numbered as described in the text.

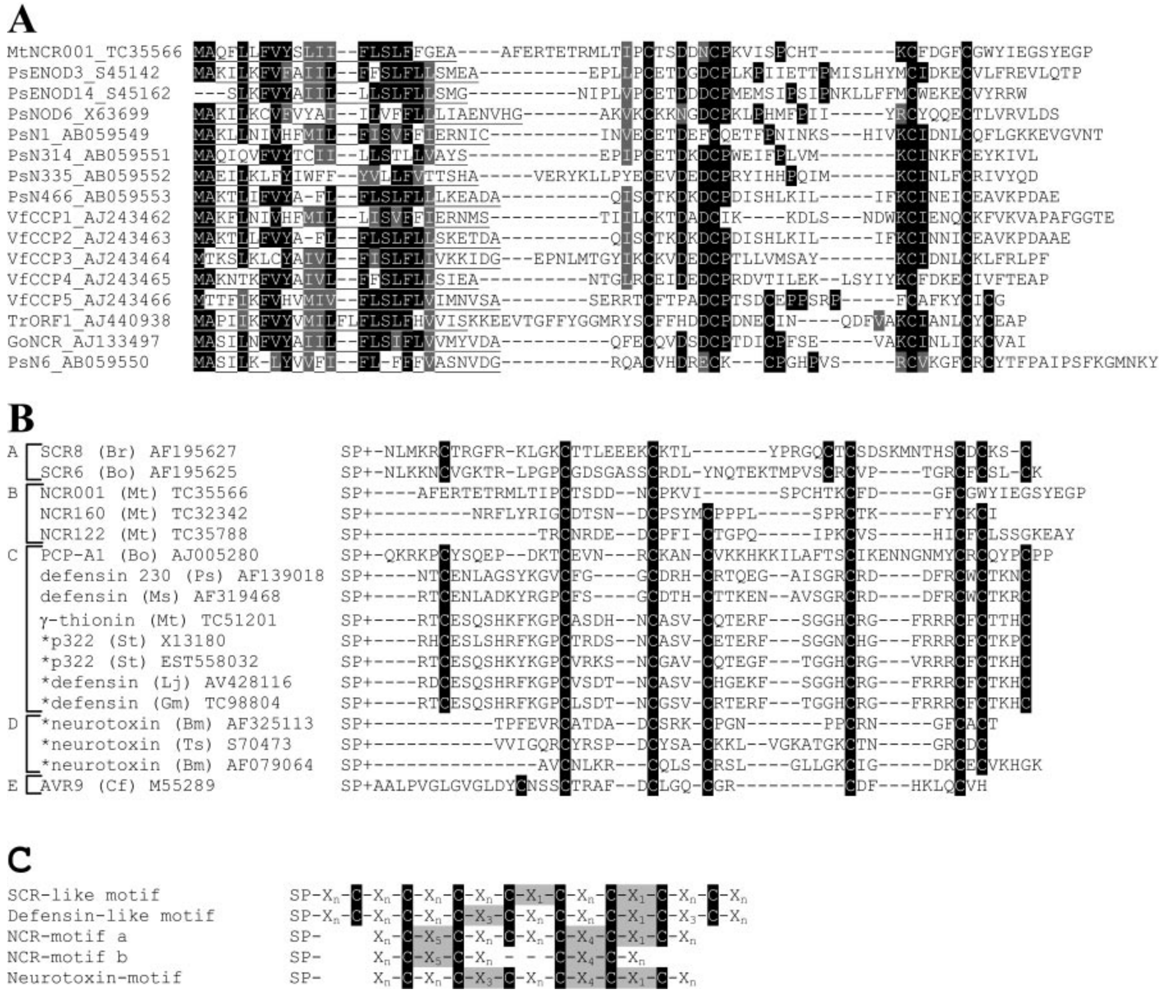

Figure 2. NCR homologs and similar polypeptides in other species. A, NCR homologs in the galegoid group of legumes. The signal peptides and the conserved residues are indicated as in Figure 1. Mt, *M. truncatula*; Ps, pea; Vf, broad bean; Go, *G. orientalis*; Tr, white clover. B, Proteins with similar structure to NCRs. Sequences indicated with asterisks were found in BLAST searches with NCR polypeptides as query. Accession numbers are indicated. The conserved Cys are shaded. SP, Signal. A, Self-incompatibility (SCR) proteins; B, NCR proteins; C, pollen coat proteins (PCPs), defensin proteins, γ-thionin proteins, and p322 proteins (the latter are incorrectly designated in databases as proteinase inhibitors); D, scorpion (*Buthus martensii* and *Thyus serrulatus*) neurotoxins; E, avirulence protein Avr9. Ser proteinase inhibitors were omitted from this alignment because of their large diversity (Laskowski and Kato, 1980). Br, *Brassica rapa*; Bo, *Brassica oleracea*; Mt, *M. truncatula*; Bm, *Buthus martensii*; Ts, *Tityus serrulatus*; Ps, pea; Ms, alfalfa; Gm, soybean; Lj, *L. japonicus*; St, *Solanum tuberosum*; Cv, *Cladosporium fulvum*. C, Cys motifs conserved in SCR-like, defensin-like, NCR-like, and neurotoxin polypeptides. Conserved Cys are in black shading, and conserved spacings are in gray shading.

Kardailsky et al., 1993; Kato et al., 2002), five from broad bean (Frühling et al., 2000), one from white clover, and one from *G. orientalis* contained the conserved signal peptide region and the Cys motifs (Fig. 2A). Pea and broad bean apparently also possess multiple members of the *NCR* family, and a global transcriptome approach would probably identify a similarly high complexity of the *NCR* family; however, so far no EST databases are available for these plants. In the TIGR EST databases containing 14 plant species including Arabidopsis, soybean, and *L. japonicus*, no more NCR polypeptides were found.

Certain known proteins both inside and outside the plant kingdom displayed structural resemblance to the NCR polypeptides, although at the primary sequence level the homology was very weak or absent (Fig. 2B). They were pollen determinants of self-incompatibility (SCR), pollen coat proteins (PCPs; Schopfer et al., 1999; Vanoosthuyse et al., 2001), defensin and γ-thionin antimicrobial peptides (Broekaert et al., 1995; Zasloff, 2002), Ser proteinase inhibitors (Laskowski and Kato, 1980), scorpion neurotoxins (Bontems et al., 1991; Froy and Gurevitz, 1998), and the fungal avirulence proteins Avr2, Avr4, and Avr9 (van Kan et al., 1991). The resemblance to NCRs was based on the small size (60–100 amino acids), the presence of a signal peptide, and conserved Cys motifs. By comparison of the Cys clusters in these polypeptides, one can discern motifs that show some common features but also classify them in different families (Fig. 2C).

For 37 *NCR* genes, we identified ESTs that carried an unspliced intron at a conserved position between the first and second nucleotide of a codon preceding the first Cys codon with a few triplets (see supplemental data at www.plantphysiol.org). Therefore, the first exon corresponded roughly to the signal peptide, whereas the second one corresponded to the mature part of the polypeptide. This intron position was also conserved in the pea and broad bean *NCR* homologous genes (Kardailsky et al., 1993; Frühling et al., 2000), but more strikingly, a similar organization was found in the plant defensin, *SCR* and *PCP* genes, and even in the scorpion toxin genes (Froy and Gurevitz, 1998; Vanoosthuyse et al., 2001; comparison of genomic and cDNA sequences for the Arabidopsis defensins AMP1 [accession nos. AC025808 and AY114038], AFP4 [accession nos. AB017065 and AY063779], and AFP4 like [accession nos. AB017065 and NM_123810]).

The Predicted Signal Peptide Targets the NCR Proteins in the Secretory Pathway

To test whether the predicted signal peptides target the NCR polypeptides in the secretory pathway, green fluorescent protein (GFP) fusions were made and transformed transiently in onion (*Allium cepa*) epidermal cells. Four different constructs were expressed from the constitutive cauliflower mosaic virus 35S promoter: a control expressing *mGFP5*, two constructs where the 3′ end of the full-length *NCR001* or *NCR084* open reading frames was fused in frame to *mGFP5*, and a fourth one carrying a NCR084 signal peptide-*mGFP5* fusion. As described by Scott et al. (1999), cells transformed with the control displayed GFP localization in the cytoplasm, in transvacuolar strands, and in the nucleus but not in the vacuole (Fig. 3, A–C). Location of the GFP signal with the three other constructs was similar to each other and distinct of the control. The NCR fusions targeted GFP to the cortical ER (Fig. 3D) and to the ER surrounding the nucleus, but the nucleus was devoid of signal (Fig. 3F), indicating that the signal peptide is functional and directs the protein to the secretory pathway. Also in the case of the NCR-GFP fusions, the vacuoles were devoid of signal (Fig. 3E). To determine whether GFP was excreted and localized in the cell wall, transformed cells were plasmolyzed separating the cytoplasm from the cell wall, but no clear cell wall-associated GFP signal was detected. Possibly, this is due to a lack in sensitivity because detection of GFP in cell walls is problematic (Scott et al., 1999), and the NCR-GFP fusions resulted in relatively weak fluorescence. However, the onion epidermal cell is a heterologous system that may lack functions present in nodule cells for proper expression and localization of the NCR polypeptides. Thus, the final destination of the NCR polypeptides has still to be determined.

The *NCR* Genes Are Nodule Specific But Have Diverse Expression Profiles during Organogenesis

Macroarrays of *M. truncatula* ESTs (Favery et al., 2002) carrying 14 distinct members of the *NCR* family were hybridized with cDNA prepared from roots and nodules at different developmental stages, from spontaneous nodules formed on alfalfa cv Sitel in the absence of a nitrogen source and rhizobia and from developmentally arrested nodules, induced by different nodulation mutants of *Sinorhizobium meliloti* (the symbiont of *M. truncatula*) or formed on mutant plants (see "Materials and Methods"). The results of the hybridization data are presented in a heat map obtained by hierarchical cluster analysis (Fig. 4A). The data are shown for the 14 *NCR* genes, a CaM-like gene, an *SPP* gene, and selected nodule-specific genes (*enod2*, *enod40*, *enod20*, *nodulin26*, and leghemoglobin genes). Expression of all the 14 *NCR* genes was nodule specific, exhibiting high expression in nodules and no expression in roots, except *NCR009*, *NCR108*, and *NCR113* showing weak expression in root tips. None of the *NCR* genes were expressed in

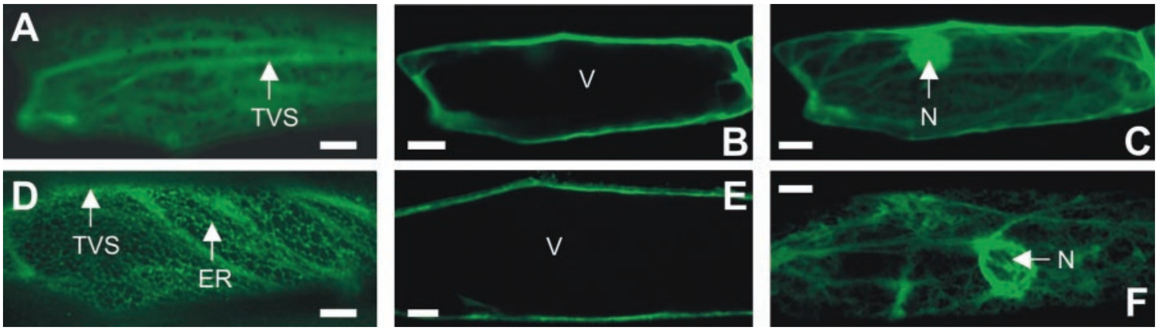

Figure 3. Subcellular localization of GFP fusion proteins in onion epidermal cells. A to C, Confocal sections of the GFP control. D to F, Confocal sections of the NCR084-GFP fusion. Sections were through the cortex (A and D), the vacuole (B and E), and the nucleus (C and F). TVS, Trans-vacuolar strands, containing cytosol and endoplasmic reticulum (ER); V, vacuole; N, nucleus.

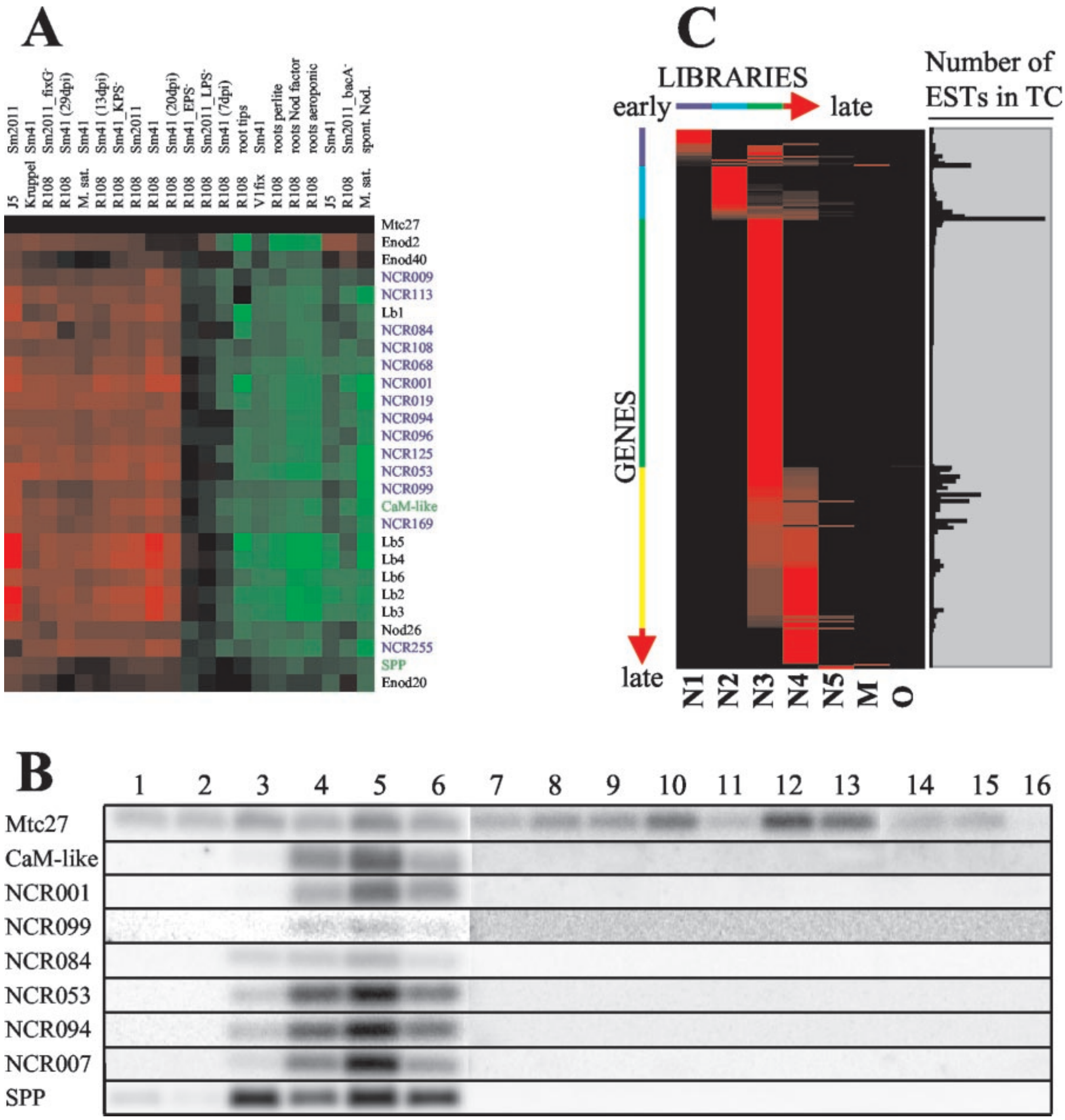

Figure 4. Expression profiles of *NCR* genes in *M. truncatula*. A, Heat map of macroarray hybridizations. *NCR* (in blue), nodulin (in black), and CaM-like and signal peptide peptidase (SPP; in green) genes are presented in the rows and hybridization experiments in the columns indicating the plant line, the bacterial strain, and the age of the nodules between brackets, if different from 3 weeks after inoculation. Black corresponds to mean level of transcript accumulation, whereas different intensities from black to red or to green indicate increasing or decreasing expression levels. B, Expression analysis by reverse transcriptase (RT)-PCR for the constitutive gene *Mtc27*, a CaM-like gene (TC51594), the nodule-expressed *SPP* genes (amplified by common primers), and the *NCR* genes *001*, *099*, *084*, *053*, *094*, and *007*. Lanes are amplifications of cDNA from: 1, roots; 2, roots treated with Nod factor; 3, Sm41-induced nodule primordia 7 d postinoculation (dpi); 4, nodules at 13 dpi; 5, nodules at 20 dpi; 6, nodules at 29 dpi; 7, cotyledons; 8, hypocotyls; 9, petioles; 10, stems; 11, apical meristems; 12, young leaves; 13, mature leaves; 14, closed flowers; 15, open flowers; 16, water control. C, Heat map of in silico expression. Red shading means relative fraction of ESTs found in the different libraries, and the histogram indicates the total number of ESTs for each *NCR* gene. The libraries and genes are ordered according to the nodule developmental stage from which they were prepared or in which they were expressed, respectively. Blue, Very early; light blue, early; green, medium; yellow, late; and red, very late. The libraries were: N1, MtBB; N2, R108Mt; N3, GVN; N4, GVSN; N5, nodulated root; M, MtBC; O, other than nodule libraries.

the bacterium-free, spontaneous nodules, except *NCR108*. The different *NCR* genes could be further distinguished by subtle differences in their expression patterns in nodules. *NCR053*, *NCR084*, *NCR094*, and *NCR096* were induced already at 7 dpi, whereas expression of the others was detected later (13 dpi).

Furthermore, the *NCR* genes expressed differently in nodule-like structures arrested early in development and induced by *S. meliloti* mutants affected in the production of exopolysaccharides (EPSs; *exoB*), lipopolysaccharides (LPS; *lpsB*), and the *bacA* mutant (no bacteroid development). In the *bacA* nodules only, *NCR084* and *NCR096* were induced. All *NCR* genes were activated in the nodule-like structures formed by the EPS and LPS mutants but at distinct attenuated levels as compared with wild type. Therefore, it can be concluded that although the tested *NCR* genes are specifically induced during nodule development, they have different patterns of expression during this process.

To confirm the expression patterns, RT-PCR assays were carried out. Specific PCR primer sets for *NCR001*, *NCR007*, *NCR053*, *NCR084*, *NCR094*, and *NCR099* were used in amplifications of cDNA from nodules at different developmental stages and from other organs (Fig. 4B). These experiments confirmed nodule-induced expression of the tested *NCR* genes and that *NCR084*, *NCR053*, *NCR094*, and *NCR007* were activated earlier (at 7 dpi) than *NCR001* or *NCR099* (at 13 dpi), as was observed in the macroarray experiments. The RT-PCR experiments further demonstrated that the tested *NCR* genes are nodule specific and not expressed in any other tested plant tissues.

To obtain expression data for all the 311 *NCR* members, the expression reports were downloaded (www.tigr.org/tdb/tgi/) and converted to a table format (see supplemental data at www.plantphysiol.org). The data were normalized and ordered as described in "Materials and Methods," and a heat map was generated to graphically display the data in an easily interpretable format (Fig. 4C). This in silico expression data proved that each member of the *NCR* family was nodule specific (all ESTs came from nodule libraries) except for *NCR122*, which was expressed in both nodules and mycorrhiza and for *NCR218* expressed only in mycorrhiza. The *NCR* genes could be classified according to the origin of the nodule cDNA libraries. The TIGR MtGI contains eight different nodule libraries prepared from different developmental stages, and the ESTs coding for the *NCR* family members were found in five of them (Fig. 4C). Four libraries (N1–N4, Fig. 4C) could clearly be ordered according to the nodule developmental stages: N1, nodule primordia, N2, young nodules; N3, mature nodules; and N4, senescent nodules (www.tigr.org/tdb/tgi/). N5 was made of pooled materials of mixed developmental stages. Ordering the *NCR* members according to the libraries (Fig. 4C) resulted in five *NCR* clusters: a very early (blue), an early (light blue), a medium (green), a late (yellow), and a very late (red) cluster. Although one should take care of the significance of such a classification for individual *NCR* members, at least it demonstrates that *NCR* genes may be expressed at different time points in the nodule development. Expression levels of the different *NCR*s ranged from one EST (106 singletons) to 113 ESTs for *NCR001* corresponding to very high expression comparable with that of the leghemoglobin *Lb1* gene with 263 EST hits. The contribution of the *NCR* family to the nodule transcriptome was extremely high, calculated to be 4.6% of the total mRNA population (1,414 *NCR* ESTs in a total of 30,707 nodule ESTs) and more than 2.5-fold higher than the contribution of all leghemoglobin genes together (558 ESTs or 1.8%; see supplemental data at www.plantphysiol.org).

Two members of the *NCR* family exhibiting different patterns in the macroarray and RT-PCR experiments (the early gene *NCR084* and the late *NCR001*) were chosen for analysis of their spatial expression profiles in nodules by in situ hybridization. The leghemoglobin *Lb1* was used as a control that expressed in the nitrogen-fixing zone (Fig. 5B). Transcripts of *NCR001* were detected in the nitrogen-fixing zone (Fig. 5C), and their accumulation started at the boundary of interzone II–III (where amyloplasts start to accumulate) and zone III (Fig. 5E). In contrast, the expression of *NCR084* was in the interzone II–III but started already in the older cell layers of zone II adjacent to interzone II–III (Fig. 5, A and D). *NCR084* transcripts were undetectable in zone III. Thus, *NCR001* and *NCR084* showed different patterns of expression, with *NCR084* expressing in the younger differentiating nodule cells, whereas *NCR001* expressed in the older nitrogen-fixing cells. These results were in good correlation with the different temporal expression patterns of these two genes.

In conclusion, the expression patterns of the *NCR* genes as determined by macroarrays, RT-PCR, in silico expression, and in situ hybridization revealed that all *NCR* genes were involved in nodulation, likely acting at different developmental stages and in different tissues and nodule cell types.

Genes Coregulated with the *NCR* Genes

Transcriptome approaches for assigning functions to unknown genes rely on the assumption that genes coregulated under a range of conditions are involved in similar functions or in the same pathway (Young, 2000). In the macroarray experiments, we identified two novel genes that were coregulated with the *NCR* genes (Fig. 4A). One of them encoded a CaM-like protein, also recently described by Fedorova et al. (2002). The other encoded an SPP (Fig. 6). As shown by RT-PCR, the CaM-like gene and the *NCR*s were exclusively expressed during nodule development (Fig. 4B). The *SPP* gene showed a basal expression in roots and was strongly induced during nodule development (Fig. 4B). In silico analysis of these nodule-enhanced CaM-like and *SPP* genes in the TIGR MtGI confirmed their nodule-specific expression. In the case of the CaM-like gene, corresponding

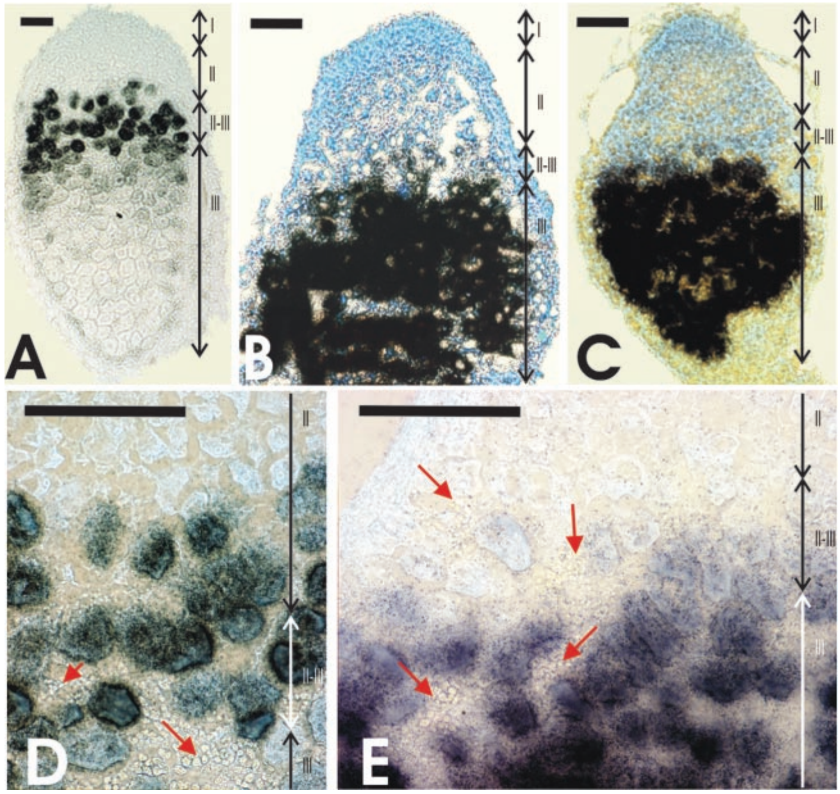

Figure 5. Localization of *NCR001* and *NCR084* transcripts in *M. truncatula* nodules. A and D, Hybridization of *M. truncatula* nodule sections with the *NCR084* antisense probe. B, Hybridization of the antisense leghemoglobin *Lb1* probe. C and E, Hybridization of the antisense *NCR001* probe. Hybridization signals are observed as black dots. The different nodule zones in the sections are indicated with arrows. The red arrows point to amyloplasts in D and E. Bars = 200 μM.

to TC51594, all the 16 ESTs derived from nodules. Moreover, six additional nodule-specific homologs of the CaM-like gene were identified (Fedorova et al., 2002). The *SPP* gene (TC44385 represented by six nodule ESTs of seven) was also part of a small gene family composed of three TCs including, another nodule-specific *SPP* gene, TC44387 (six nodule ESTs of seven) and a ubiquitous *SPP* gene, TC52930. TC44385 and TC44387 are identical in sequence except for an insertion in TC44385 (Fig. 6). Thus, these TCs could correspond to distinct genes or to alternatively spliced transcripts of the same gene. RT-PCR analysis distinguishing the two transcripts showed that TC44385, containing the insertion, was nodule specific (absent in roots), whereas the shorter transcripts (TC44387) were detectable also in roots, albeit their levels increased significantly during nodule development (data not shown).

CaMs are cytosolic Ca^{2+}-binding proteins involved in Ca^{2+} signaling (Zielinski, 1998). The CaM sequences and structures are highly conserved in all eukaryotes. The nodule-specific CaM-like proteins displayed some surprising and atypical features (Fedorova et al., 2002). They have an N-terminal extension that is absent in CaM. Although there is no experimental evidence for translation initiation at the first Met, the presence of a signal peptide sequence predicted by SignalP and its homology to the signal peptide of known nodulins suggest that the nodule-specific CaM-like proteins are secreted. Therefore, they are not only coregulated with NCRs, but they could also be colocalized. Like the NCRs, this type of CaM-like proteins was only found in the *M. truncatula* EST database, and we could not identify putative orthologs in EST databases of other plants, including other legumes.

SPP is a presenilin-type aspartic protease that catalyzes intramembrane cleavage of certain signal peptides (Weihofen et al., 2002). The *M. truncatula* SPP homologs identified here are highly homologous to the human SPP and contain all the conserved motifs, including the two aspartic acids in the protease active site (Fig. 6). Because the two *SPP* genes were coregulated with the *NCR* family, it is possible that the signal peptides of the NCR polypeptides are further cleaved by these nodule-induced SPPs, probably producing a conserved abundant oligopeptide in the nodule cells.

DISCUSSION

Large Gene Families

Here, we identified the extremely large, nodule-specific *NCR* family composed of more than 300 genes that might violate the common view on "one (or few genes) for one biological function." Although the NCR polypeptides are highly divergent, they clearly belong to the same class because: (a) All polypeptides have a relatively well conserved signal

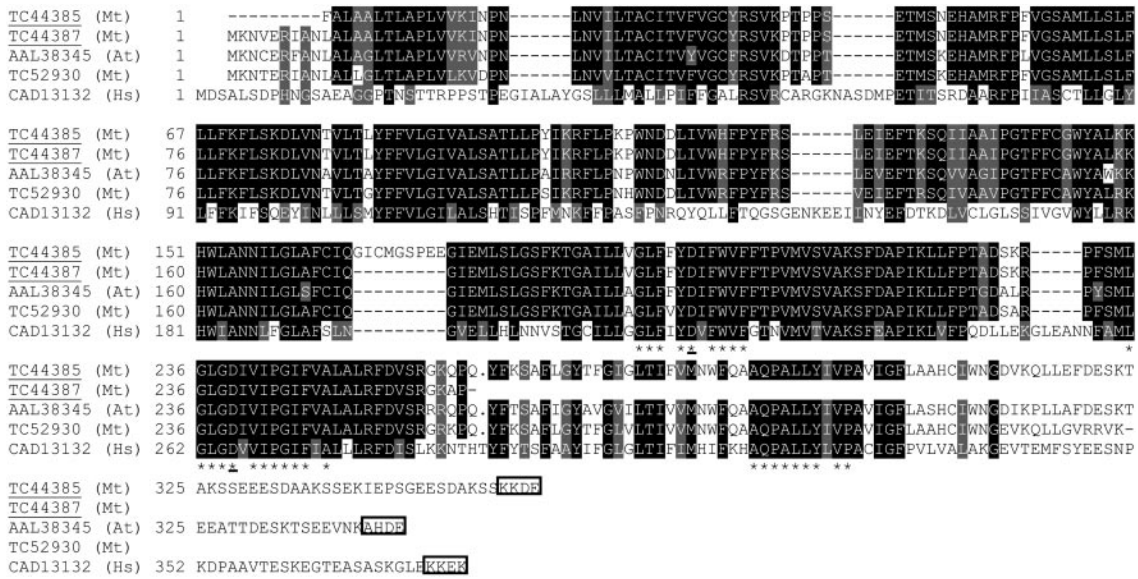

Figure 6. Alignment of SPPs. Underlined accession numbers correspond to the nodule specific TCs. The sequence of TC44385 is partial at the N terminus and that of TC44387 and TC52930 at the C terminus. The most conserved residues in the SPP family (Weihofen et al., 2002) are indicated with an asterisk, and the two aspartic acids of the active site are marked with an underlined asterisk. ER retention signals are boxed. Conserved residues are indicated as in Figure 1. Mt, *M. truncatula*; At, Arabidopsis; Hs, human (*Homo sapiens*).

peptide; (b) all of them are very small; (c) their Cys motif has unique and characteristic signatures; (d) an intron is present at conserved position; and (e) they are all nodule-specific; therefore, they are most likely functionally related.

What can be the biological meaning of this large gene family with high diversity? Several examples of large multigene families are described. In humans, the repertoire of olfactory receptors consists of about a thousand different genes, encoding highly variable seven-transmembrane receptors providing the capacity to discriminate a large number of scents (Buck and Axel, 1991). Other examples are the defensins (Schutte et al., 2002) and immunoglobulins (Cook and Tomlinson, 1995) and in plants, the receptor-like kinase genes involved in different plant signaling processes (Shiu and Bleecker, 2001), resistance genes conferring resistance to pathogen challenges by recognizing elicitors (Bergelson et al., 2001), and the SCR- and PCP-like families (Vanoosthuyse et al., 2001). A common feature of these multigene families is their involvement in recognition events, and it is not unlikely that NCR polypeptides might do so as well.

In numerous *NCR* genes, we identified an intron at a conserved position. Similar exon-intron organization was found in the *SCR* and *PCP* genes in *Brassica* spp. and the *SCR*-like, *PCP*-like, and defensin genes in Arabidopsis (Doughty et al., 1998; Schopfer et al., 1999; Vanoosthuyse et al., 2001; P. Mergaert, unpublished data). All these genes encoded similarly small proteins that were also rich in Cys. Thus, it is possible that these different gene families have a common origin. However, the sequence divergence among them is too high to assign a role for the NCR polypeptides.

Subcellular Localization of the NCR Polypeptides

Although the present results do not provide a definite answer, all the current arguments point toward an extracellular localization of the NCR polypeptides. SignalP identified the presence of a typical signal peptide that, as we confirmed experimentally for two family members, targets the NCR polypeptides to the secretory pathway. Proteins entering the secretory pathway can have any of the four possible final destinations in the nodule cells: the ER, the vacuole, the symbiosome, or the extracellular space (Verma and Hong, 1996; Neuhaus and Rogers, 1998). The first three destinations require, in addition to the signal peptide, the presence of specific address tags on the protein. In the absence of such tag, the protein follows the default pathway and is targeted to the extracellular space. In the NCR polypeptides, no specific address tags were present, nor were clear conserved peptide motifs that could serve as address tags for these polypeptides. This suggests that the NCR polypeptides follow the default pathway and, thus, are destined outside the cell, similar to the related SCR, defensins, Avr proteins, scorpion toxins, and proteinase inhibitors.

An Additional Role for the Signal Peptides?

Signal peptides are generally thought of as simple address tags directing proteins to the ER that after their cleavage end up in the garbage can of the cells. Despite a common structure, a hydrophobic domain flanked with polar regions, signal peptides display high sequence variations. Thus, it is puzzling as to why the signal peptides in the NCR family were so highly conserved. Could they play an additional role? Recent findings demonstrated that certain signal peptide fragments, generated by SPP, might have regulatory roles e.g. by binding to MHC class I molecules or to CaM (Martoglio and Dobberstein, 1998). During nodule formation, *M. truncatula* homologs of *SPP* were coregulated with the *NCR* genes. Thus, a very attractive hypothesis would be that the highly conserved NCR signal peptides are further processed by the nodule-specific SPP. Resulting oligopeptides could have an independent and most likely unique signaling function in the different nodule cell types.

Biological Role of the *NCR* Gene Family

The expression of the *NCR* gene family was restricted to developing and mature nodules, except for two members that were also expressed during mycorrhization, suggesting that their functions are related to symbiosis and principally to nodulation. The specific temporal and spatial expression profiles of the tested *NCR* genes may indicate that they have distinct roles or specializations in different nodule developmental zones and explain, at least partly, why the family is so large. The *NCR* transcripts are extremely abundant in nodules indicating that the encoded proteins are needed in high quantities. Thus, another reason for the amplification of the *NCR* family in the genome might be a gene dosage effect assuring a high expression level. The extraordinarily high expression level indicates that the family might have an extremely important role in nodule development or functioning.

Members of the *NCR* gene family have been found in the legumes belonging to the galegoid group (Fig. 7) where the expression was also nodule specific. No *NCR* members were found in nonlegumes and, more surprisingly, in the legumes *L. japonicus* or soybean. Thus, the *NCR* family might be specific for the galegoid group of legumes, and other legumes might have evolved different strategies for the function(s) accomplished by the *NCR* family. Alternatively, the role of the NCR family might be linked to the formation of indeterminate nodules by legumes of the galegoid group, whereas *L. japonicus* and soybean develop determinate nodules (Fig. 7). It will be of interest to analyze nodule EST libraries of other legumes forming determinate or indeterminate nodules (e.g. bean [determinate], *Lupinus albus* [indeterminate], or *Sesbania rostrata* capable to form both type of nodules depending on the physiological conditions; Fernández-López et al., 1998).

Because two members of the gene family were expressed in endomycorrhiza, one could speculate

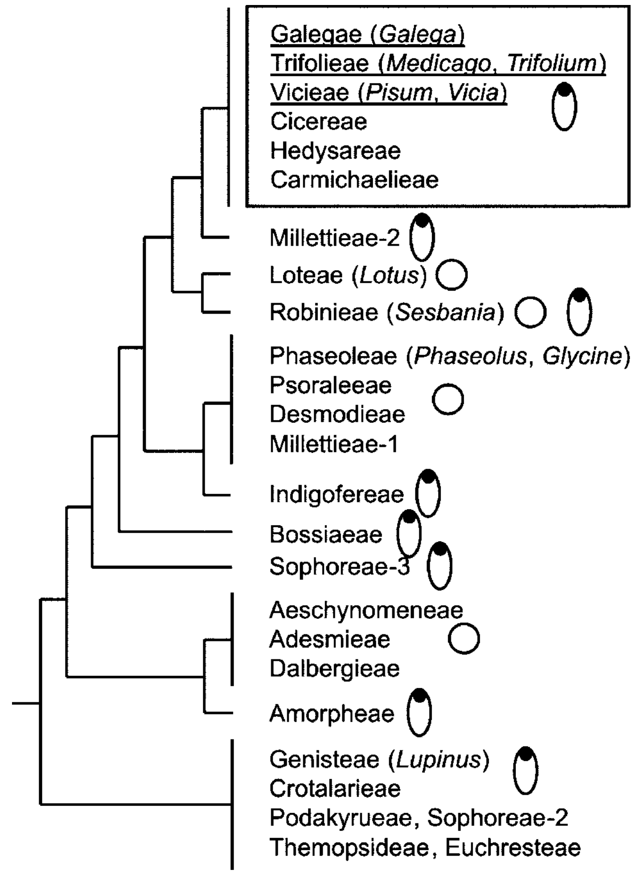

Figure 7. Phylogeny of Leguminosae based on the *rbcL* sequence. The tree is adopted from Doyle (1998), and only the branch of the *Papilionoideae* subfamily relevant to the discussion is shown. The galegoid clade is boxed, and the genera, where *NCR* family members were found, are underlined. The nodule types are symbolized by a circle for the round-shaped determinate nodules and by an oval form for the meristem containing elongated indeterminate nodules.

that an ancient *NCR* gene involved in the endomycorrhizal symbiosis has been recruited and then duplicated multiple times by the progenitor of galegoid legumes for the nitrogen-fixing symbiosis. *Enod40*, *Enod12*, or *Nod26* are other nodulin genes that may have a similar evolutionary history (Albrecht et al., 1999). Moreover, a large number of plant mutants are affected in both types of endosymbiosis (Albrecht et al., 1999; Kistner and Parniske, 2002). Thus, there exists a large overlap between the mechanisms leading to these symbioses.

Biochemical Role of NCR Polypeptides

Members of the NCR family were first described in pea (Scheres et al., 1990; Kardailsky et al., 1993) and the authors, based on the presence of the Cys residues, suggested a role for these proteins in metal binding and transport, providing the bacteroids with the necessary metals for nitrogenase functioning. However, no evidence has ever been provided for such a function. Metal-binding proteins or peptides are cytosolic and bind metals via a free thiolate (R-S^{-}) group (Robinson et al., 1993). However, the NCR polypeptides are targeted into the secretory pathway. Therefore, the Cys are probably involved in the formation of disulfide bridges and unable to complex metals. All of the proteins that exhibit structural resemblance to NCRs have a globular structure consolidated by one to four disulfide bonds (Laskowski and Kato, 1980; Broekaert et al., 1995; Froy and Gurevitz, 1998). The NCR proteins might also be globular proteins stabilized by two or three disulfide bridges. Thus, involvement of NCRs in metal binding and transport is rather unlikely.

The NCR polypeptides might act as antimicrobial defensins or diffusible signaling molecules assuring cell-to-cell communication. If NCRs were antimicrobial peptides, their function could be avoiding opportunistic infections by other soil microorganisms during nodule formation or confining the rhizobia inside the nodules. Their extreme sequence variations could provide a potent cocktail of antimicrobial peptides with a broad spectrum. Alternatively, because polypeptides are an emerging type of signals in plants (Lindsey et al., 2002), NCR polypeptides could assure communication between plant cells or between plant cells and bacteria. Nodule formation and growth of mature nodules involves differentiation of plant and bacterial cells requiring a strong coordination relying on cell-to-cell signaling. The tested *NCR* transcripts were localized in different cell layers and zones. It is not unlikely that development of the different stages of symbiotic cells could be coupled to expression of different sets of *NCR*s. The NCR polypeptides, acting as signals, could mediate consecutive differentiation events both in the plant cells and in the symbiosomes. In addition, the nodule-specific CaM-like proteins could be partners of NCR polypeptides in these functions.

Whatever the molecular role of the NCR polypeptides is, the high number of genes and the molecular diversity within them indicate that there exists in legume nodules an unanticipated complexity of interactions between plant cells, plant cells and rhizobia or plant cells and other rhizosphere microorganisms.

MATERIALS AND METHODS

Sequence Analysis

BLAST searches were made at TIGR (www.tigr.org/tdb/tgi/) or at the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/BLAST/). The NCR family was identified by repetitive searches at the TIGR MtGI (Release 4.0) with BLASTn and TBLASTn, using *NCR001* as the first sequence and repeating the searches, each time with novel homologous sequences, until no further family members were found. Singletons or TCs containing introns were recognized by BLASTn alignments with the corresponding singletons or TCs lacking the intron. Sequence alignments were made with ClustalW at www.ebi.ac.uk/clustalw/. Signal peptides and putative cleavage sites were predicted by SignalP (www.cbs.dtu.dk/services/).

Transformation and Visualization of GFP Fusion Proteins in Onion (*Allium cepa*) Epidermal Cells

PCR fragments, corresponding to the full-length open reading frames of *NCR001* and *NCR084* and to the signal peptide part of *NCR084*, were cloned in the pCAMBIA1302 vector (www.cambia.org). Transformation of onion epidermal cells was done according to Scott et al. (1999). Plasmolysis of the onion cells was done as described by Scott et al. (1999).

Plant Lines, *Sinorhizobium meliloti* strains, and Growth Conditions

Medicago truncatula lines R108 and Jemalong J5 and alfalfa (*Medicago sativa*) subsp. *sativa* cv Sitel were inoculated with *S. meliloti* to obtain wild-type ($Nod^{+}Fix^{+}$) nodules. Spontaneous nodules were obtained on selected lines of alfalfa cv Sitel grown in nitrogen-free medium in the absence of rhizobia. $Nod^{+}Fix^{-}$ nodules were obtained from transgenic R108 plants, expressing the antisense construct of the "Krüppel"-like *Mszpt2-1* gene (Frugier et al., 2000) and from the V1 somaclonal mutant forming nodule primordia but no nodules (gift from Pascal Ratet, Institut des Sciences du Végétal-Centre National de la Recherche Scientifique, Gif-sur-Yvette, France). The wild-type *S. meliloti* strains were Sm41 and Sm2011 (Sm2011 forms nitrogen-fixing nodules on R108, J5, and Sitel; Sm41 forms nitrogen-fixing nodules on R108 and Sitel but only nodule primordia on J5). Nodule primordia developed on R108 by AK631 corresponding to Sm41*exoB* lacking EPS production; Sm41*lpsZ* mutated for KPS production (PP699; Putnoky et al., 1990); Sm2011*lpsB*, an LPS synthesis mutant (Sm6963; Niehaus et al., 1998); and Sm1021*bacA* impaired in infection and bacteroid differentiation (Sm8368; Glazebrook et al., 1993). Fix^{-} nodules were obtained on R108 by Sm2011*fixG* impaired in the nitrogen fixation complex (GMI394; Kahn et al., 1989). Bacterial and plant cultures, plant inoculation, and nodulation were done as described at www.isv.cnrs-gif.fr/embo2/manuels/index.html. For the kinetics of nodule development, nodules were collected at 7, 13, 20, and 29 dpi; otherwise, nodules were harvested 3 weeks after inoculation or nitrogen starvation. Root material was obtained from plants grown under axenic but otherwise identical conditions as nodulated plants.

Expression Analysis

The macroarrays, protocols for RNA extractions, labeling, hybridization and quantification of hybridization signals were as described at www.isv.cnrs-gif.fr/embo2/manuels/index.html. Hybridizations were repeated twice. The data points for the genes of interest (Fig. 4A) were extracted from the raw data of the hybridization experiments. The experiments are normalized relative to the constitutive *Mtc27* expression (Györgyey et al., 1991). Data were treated with the Cluster and TreeView software (http://rana.lbl.gov/EisenSoftware.htm; Eisen et al., 1998). Cluster was used to "log transform" the data, "mean center" the genes, and for hierarchical clustering of the genes and experiments. TreeView was used to visualize the data.

The data set for the in silico expression analysis of the *NCR* family was downloaded from www.tigr.org/tdb/tgi/using the expression reports for TCs and the EST reports for singletons. The EST counts per cDNA library were converted to an Excel table (Microsoft, Redmond, WA; see supplemental data at www.plantphysiol.org). These counts were used as raw expression data, analogous to raw array hybridization results. The counts were normalized for the total number of ESTs per library and for the total number of ESTs per TC. The nodule cDNA libraries were ordered according to the developmental stage from which they were prepared (very early, MtBB, TIGR no. 5519; early, R108Mt, TIGR no. 2764; late, GVN, TIGR no. T1617; and very late, GVSN, TIGR no. T10109). The library of nodulated roots (TIGR no. 4047) and the MtBC library (TIGR no. 5520) prepared from mycorrhizal roots were also included separately in the analysis. All other libraries were treated together. Then, the genes were ordered according to their distribution in the ordered libraries. The TreeView software was used for visualization of the resulting expression patterns.

RT-PCR experiments were done as described by Kevei et al. (2002). The in situ hybridizations were done as described by de Almeida Engler et al. (2001) and Kevei et al. (2002).

Distribution of Materials

Upon request, all novel materials described in this publication will be made available in a timely manner for noncommercial research purposes, subject to the requisite permission from any third party owners of all or parts of the material. Obtaining any permission will be the responsibility of the requestor.

ACKNOWLEDGMENTS

We thank Carol Vance (University of Minnesota, St. Paul) for sharing results before publication; our colleagues Pascal Ratet, Hanh Trihn, and Martin Crespi (Institut des Sciences du Végétal-Centre National de la Recherche Scientifique, Gif-sur-Yvette, France) for *M. truncatula* seeds; Miguel Redondo and Gabriella Jahni (Institut des Sciences du Végétal-Centre National de la Recherche Scientifique, Gif-sur-Yvette, France) for help with informatics; and Raimundo Villarroel and Christian Chaparro Egaña (University of Gent, Belgium) for help in the fabrication of macroarrays.

Received November 22, 2002; returned for revision February 9, 2003; accepted February 9, 2003.

LITERATURE CITED

Albrecht C, Geurts R, Bisseling T (1999) Legume nodulation and mycorrhizae formation; two extremes in host specificity meet. EMBO J **18:** 281–288

Bergelson J, Kreitman M, Stahl EA, Tian D (2001) Evolutionary dynamics of plant *R*-genes. Science **292:** 2281–2285

Bontems F, Roumestand C, Gilquin B, Menez A, Toma F (1991) Refined structure of charybdotoxin: common motifs in scorpion toxins and insect defensins. Science **254:** 1521–1523

Broekaert WF, Terras FRG, Cammue BPA, Osborn RW (1995) Plant defensins: novel antimicrobial peptides as components of the host defense system. Plant Physiol **108:** 1353–1358

Buck L, Axel R (1991) A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. Cell **65:** 175–187

Cook GP, Tomlinson IM (1995) The human immunoglobulin V_H repertoire. Immune Today **16:** 237–242

Crespi M, Gálvez S (2000) Molecular mechanisms in root nodule development. J Plant Growth Regul **19:** 155–166

de Almeida Engler J, De Groodt R, Van Montagu M, Engler G (2001) *In situ* hybridization to mRNA of *Arabidopsis* tissue sections. Methods **23:** 325–334

Doughty J, Dixon S, Hiscock SJ, Willis AC, Parkin IAP, Dickinson HG (1998) PCP-A1, a defensin-like *Brassica* pollen coat protein that binds the *S* locus glycoprotein, is the product of gametophytic gene expression. Plant Cell **10:** 1333–1347

Doyle JJ (1998) Phylogenetic perspectives on nodulation: evolving views of plants and symbiotic bacteria. Trends Plant Sci **3:** 473–478

Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA **95:** 14863–14868

Endre G, Kereszt A, Kevei Z, Milhacea S, Kaló P, Kiss GB (2002) A receptor kinase gene regulating symbiotic nodule development. Nature **417:** 962–966

Favery B, Complainville A, Vinardell JM, Lecomte P, Vaubert D, Mergaert P, Kondorosi A, Kondorosi E, Crespi M, Abad P (2002) The endosymbiosis-induced genes *ENOD40* and *CCS52a* are involved in endoparasitic-nematode interactions in *Medicago truncatula*. Mol Plant-Microbe Interact **15:** 1008–1013

Fedorova M, van de Mortel J, Matsumoto PA, Cho J, Town CD, VandenBosch KA, Gantt S, Vance CP (2002) Genome-wide identification of nodule-specific transcripts in the model legume *Medicago truncatula*. Plant Physiol **130:** 519–537

Fernández-López M, Goormachtig S, Gao M, D'Haeze W, Van Montagu M, Holsters M (1998) Ethylene-mediated phenotypic plasticity in root nodule development on *Sesbania rostrata*. Proc Natl Acad Sci USA **95:** 12724–12728

Froy O, Gurevitz M (1998) Membrane potential modulators: a thread of scarlet from plants to humans. FASEB J **12:** 1793–1796

Frugier F, Poirier S, Satiat-Jeunemaître B, Kondorosi A, Crespi M (2000) A Krüppel-like zinc finger protein is involved in nitrogen-fixing root nodule organogenesis. Genes Dev **14:** 475–482

Frühling M, Albus U, Hohnjec N, Geise G, Pühler A, Perlick AM (2000) A small gene family of broad bean codes for late nodulins containing conserved cysteine clusters. Plant Sci **152:** 67–77

Glazebrook J, Ichige A, Walker GC (1993) A *Rhizobium meliloti* homolog of the *Escherichia coli* peptide-antibiotic transport protein SbmA is essential for bacteroid development. Genes Dev **7:** 1485–1497

Györgyey J, Gartner A, Németh K, Magyar Z, Hirt H, Heberle-Bors E, Dudits D (1991) Alfalfa heat shock genes are differentially expressed during somatic embryogenesis. Plant Mol Biol **16:** 999–1007

Györgyey J, Vaubert D, Jiménez-Zurdo JI, Charon C, Troussard L, Kondorosi A, Kondorosi E (2000) Analysis of *Medicago truncatula* nodule expressed sequence tags. Mol Plant-Microbe Interact **13:** 62–71

Kahn D, David M, Domergue O, Daveran M-L, Ghai J, Hirsch PR, Batut J (1989) *Rhizobium meliloti fixGHI* sequence predicts involvement of a specific cation pump in symbiotic nitrogen fixation. J Bacteriol **171:** 929–939

Kardailsky I, Yang W-C, Zalensky A, van Kammen A, Bisseling T (1993) The pea late nodulin gene PsNOD6 is homologous to the early nodulin genes PsENOD3/14 and is expressed after the leghaemoglobin genes. Plant Mol Biol **23:** 1029–1037

Kato T, Kawashima K, Miwa M, Mimura Y, Tamaoki M, Kouchi H, Suganuma N (2002) Expression of genes encoding late nodulins characterized by a putative signal peptide and conserved cysteine residues is reduced in ineffective pea nodules. Mol Plant-Microbe Interact **15:** 129–137

Kevei Z, Vinardell JM, Kiss GB, Kondorosi A, Kondorosi E (2002) Glycine-rich proteins encoded by a nodule-specific gene family are implicated in different stages of symbiotic nodule development in *Medicago*. Mol Plant-Microbe Interact **15:** 922–931

Kistner C, Parniske M (2002) Evolution of signal transduction in intracellular symbiosis. Trends Plant Sci **7:** 511–518

Laskowski M Jr, Kato I (1980) Protein inhibitors of proteinases. Annu Rev Biochem **49:** 593–626

Lindsey K, Casson S, Chilley P (2002) Peptides: new signalling molecules in plants. Trends Plant Sci **7:** 78–83

Martoglio B, Dobberstein B (1998) Signal sequences: more than just greasy peptides. Trends Cell Biol **8:** 410–415

Neuhaus J-M, Rogers JC (1998) Sorting of proteins to vacuoles in plant cells. Plant Mol Biol **38:** 127–144

Niehaus K, Lagares A, Pühler A (1998) A *Sinorhizobium meliloti* lipopolysaccharide mutant induces effective nodules on the host plant *Medicago sativa* (alfalfa) but fails to establish a symbiosis with *Medicago truncatula*. Mol Plant-Microbe Interact **9:** 906–914

Putnoky P, Petrovics G, Kereszt A, Grosskopf E, Ha DTC, Banfalvi Z, Kondorosi A (1990) *Rhizobium meliloti* lipopolysaccharide and exopolysaccharide can have the same function in the plant-bacterium interaction. J Bacteriol **172:** 5450–5458

Quackenbush J, Cho J, Lee D, Kiang F, Holt I, Karamycheva S, Parvizi B, Pertea G, Sultana R, White J (2001) The TIGR gene indices: analysis of gene transcript sequences in highly sampled eukaryotic species. Nucleic Acids Res **29:** 159–164

Robinson NJ, Tommey AM, Kuske C, Jackson PJ (1993) Plant metallothioneins. Biochem J **295:** 1–10

Schauser L, Roussis A, Stiller J, Stougaard J (1999) A plant regulator controlling development of symbiotic root nodules. Nature **402:** 191–195

Scheres B, van Engelen F, van der Knaap E, van de Wiel C, van Kammen A, Bisseling T (1990) Sequential induction of nodulin gene expression in the developing pea nodule. Plant Cell **2:** 687–700

Schopfer CR, Nasrallah ME, Nasrallah JB (1999) The male determinant of self-incompatibility in *Brassica*. Science **286:** 1697–1700

Schutte BC, Mitros JP, Bartlett JA, Walters JD, Jia HP, Welsh MJ, Casavant TL, McCray PB Jr (2002) Discovery of five conserved β-defensin gene clusters using a computational search strategy. Proc Natl Acad Sci USA **99:** 2129–2133

Scott A, Wyatt S, Tsou P-L, Robertson D, Strömgren Allen N (1999) Model system for plant cell biology: GFP imaging in living onion epidermal cells. BioTechniques **26:** 1125–1132

Shiu S-H, Bleecker AB (2001) Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. Proc Natl Acad Sci USA **98:** 10763–10768

Stracke S, Kistner C, Yoshida S, Mulder L, Sato S, Kaneko T, Tabata S, Sandal N, Stougaard J, Szczyglowski K et al. (2002) A plant receptor-like kinase required for both bacterial and fungal symbiosis. Nature **417:** 959–962

van Kan JAL, van den Ackerveken GFJM, de Wit PJGM (1991) Cloning and characterization of cDNA of avirulence gene *avr9* of the fungal pathogen *Cladosporium fulvum*, causal agent of tomato leaf mold. Mol Plant-Microbe Interact **4:** 52–59

Vanoosthuyse V, Miege C, Dumas C, Cock M (2001) Two large *Arabidopsis thaliana* gene families are homologous to the *Brassica* gene superfamily that encodes pollen coat proteins and the male component of the self-incompatibility response. Plant Mol Biol **16:** 17–34

Vasse J, de Billy F, Camut S, Truchet G (1990) Correlation between ultrastructural differentiation of bacteroids and nitrogen fixation in alfalfa nodules. J Bacteriol **172:** 4295–4306

Verma DPS, Hong Z (1996) Biogenesis of the peribacteroid membrane in root nodules. Trends Microbiol **4:** 364–368

Weihofen A, Binns K, Lemberg MK, Ashman K, Martoglio B (2002) Identification of signal peptide peptidase, a presenilin-type aspartic protease. Science **296:** 2215–2218

Young RA (2000) Biomedical discovery with DNA arrays. Cell **102:** 9–15

Zasloff M (2002) Antimicrobial peptides of multicellular organisms. Nature **415:** 389–395

Zielinski RE (1998) Calmodulin and calmodulin-binding proteins in plants. Annu Rev Plant Physiol Plant Mol Biol **49:** 697–725